1999 First Year Exam
In Class Portion

1. This is a closed book exam. No notes are permitted.
2. It is to your advantage to show your work and explain your answers. Don't erase. Draw a line through work you don't want graded.
3. Use a separate page for each problem so different people can grade different problems.
4. Put your code on each page of your exam.
5. Partial credit is given. A few complete answers are more valuable than many incomplete answers.
6. You have three hours to finish.

Problem 1

Suppose $Y|X, \beta$ is $\mathrm{N}(X\beta, \sigma^2\mathrm{I})$.

a) Choose a non-informative prior on (β, σ^2) and derive the joint posterior of (β, σ^2), and the marginal posterior distributions of β and σ^2.

b) Now suppose that X and β can be partitioned into blocks as (X_1, X_2) and (β_1, β_2) so that X_1 and β_1 correspond to "fixed" effects and X_2 and β_2 correspond to "random" effects. Specify a hierarchical linear model that includes parameters of interest for both fixed and random effects, and either derive the marginal posterior distributions for the parameters of interest or derive the conditional distributions $[\beta_1|X, Y, \beta_2, \sigma^2]$, $[\beta_2|X, Y, \beta_1, \sigma^2]$, and $[\sigma^2|X, Y, \beta_1, \beta_2]$.

Problem 2

Let (X, Y) have a bivariate Normal distribution with mean $(0, 0)$ and covariance matrix $\sigma^2\mathrm{I}$. (X, Y) can be reparametrized as (T, Z) where Z is the distance from the origin to the point (X, Y) and T is the angle between the vector (X, Y) and the x-axis:

$$T = \arctan(Y/X)$$
$$Z = (X^2 + Y^2)^{1/2}$$

a) Find the joint density of (T, Z). Please note any assumptions you make.

b) Find a sufficient statistic for the parameter σ^2 based on an i.i.d. sample $z_1, \ldots, z_n$ having the distribution of Z. You may use the factorization criterion if you wish.

Problem 3

Suppose that $X_1, \ldots, X_n$ are i.i.d. Uniform$(0, \theta)$. Let $M_n = \max(X_1, \ldots, X_n)$ and consider the test

$$\delta_c(X) = \begin{cases} 1 & \text{if } M_n \geq c, \\ 0 & \text{otherwise.} \end{cases}$$

Here $\delta_c(X) = 1$ indicates rejection of H_0 and $\delta_c(X) = 0$ indicates failure to reject H_0.

a) Compute the power function of δ_c.

b) In testing $H_0 : \theta = 1/2$ vs. $H_1 : \theta > 1/2$ what choice of c would make δ_c have size exactly 0.05?

c) How large should n be so that the δ_c specified in (b) has power 0.98 for $\theta = 3/4$?

d) If $n = 20$ and $M_n = 0.48$, what is the p-value?

e) Using a Gamma prior, $\theta \sim \text{Ga}(\alpha, \beta)$, find the posterior probability $\Pr[\theta \leq 1/2 | X_1, \ldots, X_n]$. You may leave the answer as a quantile of a distribution.

f) In the setup of (e), find the posterior expectation $\text{E}[X_{n+1} | X_1, \ldots, X_n]$.

Problem 4

A botanist is studying tree seedlings. In November of year 1 she marks a one meter square plot in the forest, finds all the tree seedlings in the plot and attaches flags to them.

When trees lose their leaves in the winter they acquire leaf scars. The botanist examines each seedling for leaf scars to determine whether or not it has lived through its first winter. She attaches red flags to those that have survived at least one winter (Old seedlings) and green flags to those that have not (New seedlings, ones that have just sprouted the preceding summer).

In November of year 2 she returns to her plot to see which seedlings have survived. Let θ_{Old} be the probability that a randomly selected Old seedling survives and θ_{New} be the probability that a randomly selected New seedling survives. Let n_{Old} and n_{New} be the numbers of Old and New seedlings respectively that the botanist flagged in year 1. Let X_{Old} and X_{New} be the numbers of Old and New seedlings respectively that have survived to year 2.

a) Write down the likelihood function for $(\theta_{\text{Old}}, \theta_{\text{New}})$. Does it form an exponential family? Write down a convenient prior and calculate the posterior.

b) Now suppose that sometime between November of year 1 and November of year 2 a group of children playing the forest removed all the flags. Of the seedlings the botanist finds in year 2, she cannot tell which were Old and which were New in year 1. In other words, instead of observing $(X_{\text{Old}}, X_{\text{New}})$, she can only observe $Y = X_{\text{Old}} + X_{\text{New}}$. Write down the likelihood function for $(\theta_{\text{Old}}, \theta_{\text{New}})$. Does it form an exponential family?

c) Under the conditions of part b), how would you calculate the posterior? (Don't calculate it; just say how you would do it.)

d) Describe a Gibbs sampling algorithm to calculate the posterior in part b). You may introduce an auxiliary variable if you wish.

Problem 5

a) A stick of length ℓ is randomly broken into three pieces. The two break locations are i.i.d. $U[0, \ell]$ random variables. What is the chance that the pieces can be put together in a triangle?

b) Suppose the outcome X of a certain chance mechanism depends on p according to $\Pr\{X = 1\} = p$, $\Pr\{X = 0\} = 1 - p$, where $0 \leq p \leq 1$. Suppose p is chosen at random, uniformly distributed over $[0, 1]$, and then that two independent outcomes X_1 and X_2 are observed. What is the unconditional correlation between X_1 and X_2?

Problem 6

Consider n systems with failure times $X_1, X_2, \ldots, X_n$, assumed to be independent and identically distributed with exponential, $\mathrm{Exp}(\lambda)$, distributions.

a) Find a method of moments estimate of λ.

b) Find a method of moments estimate of λ different from the one in a).

c) Find the m.l.e. $\hat{\lambda}$ for λ.

d) Find a method of moments estimate of the probability $\eta = \Pr[X_i \geq 1]$.

e) Find the m.l.e. $\hat{\eta}$ for η.

f) Assuming a gamma prior, $\lambda \sim \mathrm{Ga}(\alpha, \beta)$, find the posterior mean $\bar{\lambda}_n = \mathrm{E}[\lambda | X_1, \ldots, X_n]$.

1999 First Year Exam
Take Home Portion

1. This is an open book exam. You may use any resources except other people.
2. Put your code on each page of your exam.
3. There is no single right answer.
4. Partial credit is given.
5. You have twelve hours to finish.

A study is trying to establish whether there is a relationship between a given gene and allergy to tree pollen. To simplify suppose that we can capture allergy to tree pollen by a binary variables (one is either allergic or not and knows it). The genotype (what type of gene one has) can take three discrete values, usually labelled AA, Aa and aa. It is known that people with genotypes Aa and aa have the same probability of allergy to tree pollen. People with genotype AA may have a different probability of allergy.

112 randomly chosen individuals are tested for the gene and asked whether they are allergic to tree pollen. In addition, each individual is asked whether his or her oldest child is allergic to tree pollen. The children cannot be tested for genotype.

The three genotypes are randomly distributed in the population with frequencies 1/9 for AA, 4/9 for Aa, and 4/9 for aa. The following table shows conditional probabilities for a child's genotype given its parents' genotypes, where the three numbers in each cell are the probabilities of AA, Aa, and aa, respectively.

Mother's genotype			
	Father's genotype		
	AA	Aa	aa
AA	(1,0,0)	(1/2,1/2,0)	(0,1,0)
Aa	(1/2,1/2,0)	(1/4,1/2,1/4)	(0,1/2,1/2)
aa	(0,1,0)	(0,1/2,1/2)	(0,0,1)

The data are in `~michael/fyedata`. The first few lines are reproduced here:

```
aa 0 0
Aa 0 1
aa 0 0
Aa 0 0
Aa 0 0
```

The first column is the genotype of the parent. The second column is 1 if the parent is allergic to tree pollen and 0 otherwise. The third column is 1 if the child is allergic to tree pollen and 0 otherwise.
Your assignment is to report on
a) whether allergy to tree pollen is related to genotype, and
b) if so, what is the likely size of the effect.

You may choose to use all the data in one analysis, or you may analyze only the parents' data formally and use the childrens' data for informal confirmation. Write a complete report that describes the problem, your modelling, your assumptions, and your conclusion.